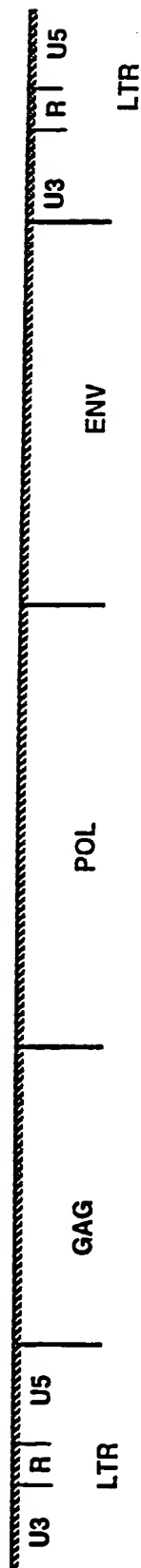
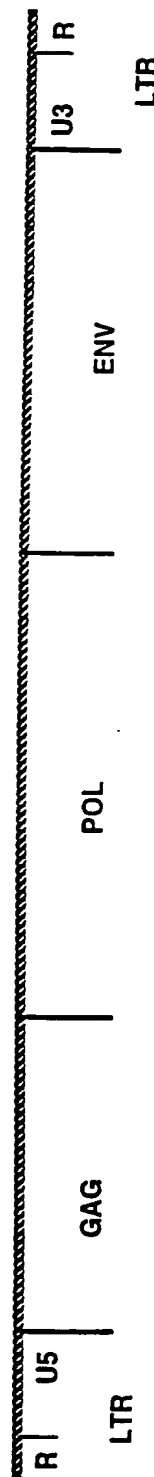


1 / 32
FIG 1

PROVIRAL DNA



GENOMIC RNA (VIRION)



2 / 32
FIG 2

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
GCTTATAGAA	GGACCCCTAG	TATGGGGTAA	TCCCCTCTGG	GAAACCAAGC	50
A Y R R	T P S M G .	S P L G	N Q A		
L I E	G P L V	W G N	P L W	E T K P	
L . K	D P .	Y G V I	P S G	K P S	
CCAGTACTC	AGCAGGAAA	ATAGAATAGG	AAACCTCACA	AGGACATACT	100
P V L	S R K N	R I G	N L T	R T Y F	
Q Y S	A G K I E .	E T S Q	G H T		
P S T Q	Q E K .	N R	K P H K	D I L	
TTCTCCCCCT	CCAGATGGCT	AGCCACTGAG	GAAGGAAAA	TACTTTTACC	150
P P L	Q M A S H .	G R K N	T F T		
F L P S	R W L A T E	E G K I	L S P		
S S P	P D G .	P L R	K E K	Y F H L	
TGCAGCTAAC	CAACAGAAAT	TACTTAAAC	CCTTCACCAA	ACCTTCCACT	200
C S .	P T E I T .	N P S P N	L P L		
A A N	Q Q K L	L K T	L H Q	T F H L	
Q L T	N R N	Y L K P	F T K	P S T	
TAGGCATTGA	TAGCACCCAT	CAGATGGCCA	AATTATTATT	TACTGGACCA	250
R H .	. H P S	D G Q	I I I	Y W T R	
G I D	S T H	Q M A K	L L F	T G P	
. A L I	A P I	R W P	N Y Y L	L D Q	
GGCCTTTTCA	AAACTATCAA	GAAGATAGTC	AGGGGCTGTG	AAGTGIGCCA	300
P F Q	N Y Q	E D S Q	G L .	S V P	
G L F K	T I K	K I V	R G C E	V C Q	
A F S	K L S R	R .	S G A V	K C A K	
AAGAAATAAT					310
K K .					
R N N					
E I					

3 / 32
FIG 2 (continued)

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
CCCTGTATCT	TTAACCTCCT	TGTTAAGTTT	GTCTCTTCCA	GAATCAAAC	50
P C I F	N L L	V K F	V S S R	I K T	
P V S	L T S L	L S L	S L P	E S K L	
L Y L	. P P	C . V C	L F Q	N Q N	
TGTAATACTA	CAAATIGTTC	TTCAAATGGA	GCACCAGATG	GAGTCCATGA	100
V K L	Q I V L	Q M E	H Q M	E S M T	
. N Y	K L F	F K W S	T R W	S P .	
C K T T	N C S	S N G	A P D G	V H D	
CTAAGATCCA	CCGTGGACCC	CTGGACCGGC	CTGCTAGCCC	ATGCTCCGAT	150
K I H	R G P	L D R P	A S P	C S D	
L R S T	V D P	W T G	L L A H	A P M	
. D P	P W T P	G P A C	. P	M L R C	
GTTAATGACA	TTGAAGGCAC	CCCTCCCGAG	GAAATCTCAA	CTGCACAACC	200
V N D I	E G T	P P E	E I S T	A Q P	
L M T	L K A P	L P R	K S Q	L H N P	
. . H	. R H	P S R G	N L N	C T T	
CCTACTATGC	CCCAATTCAG	CGGGAAGCAG	TTAGAGCGGT	CATCAGCCAA	250
L L C	P N S A	G S S	. S G	H Q P T	
Y Y A	P I Q	R E A V	R A V	I S Q	
P T M P	Q F S	G K Q	L E R S	S A N	
CCCTCCCAAC	AGCACTTGGG	TTTTCTGTGT	GAGAGGGGGG	ACTGAGAGAC	300
S P T	A L G	F S C	. E G G	L R D	
P P Q Q	H L G	F P V	E R G D	. E T	
L P N	S T W V	F L L	R G G	T E R Q	
AGGACTAGCT	GGATTTCCTA	GGCCAACGAA	GAATCCCTAA	GCCTAGCTGG	350
R T S W	I S .	A N E	E S L S	L A G	
G L A	G F P R	P T K	N P .	A . L G	
D . L	D F L	G Q R R	I P K	P S W	

4 / 32
FIG 3

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
GAAGGTGACT	GCATCCACCT	CTAAACATGG	GGCTTGCAAC	TTAGCTCACA	400
K V T	A S T S	K H G	A C N	L A H T	
R .	L H P P	L N M G	L A T	. L T	
E G D C	I H L	. T W	G L Q L	S S H	
CCCCACCAAT	CAGAGAGCTC	ACTAAAATGC	TAATTAGGCA	AAAATAGGAG	450
R P I	R E L	T K M L	I R Q	K . E	
P D Q S	E S S	L K C	. L G K	N R R	
P T N	Q R A H	. N A N	. A K	I G G	
GTAAAGAAAT	AGCCAATCAT	CTATTGCCTG	AGAGCACAGC	GGGAGGGACA	500
V K K .	P I I	Y C L	R A Q R	E G Q	
. R N	S Q S S	I A .	E H S	G R D K	
K E I	A N H	L L P E	S T A	G G T	
AGGATCGGGA	TATAAACCCA	GGCATTGCGAG	COGGCAACGG	CAACCCCTT	550
G S G	Y K P R	H S S	R Q R	Q P P L	
D R D	I N P	G I R A	G N G	N P L	
R I G I	. T Q	A F E	P A T A	T P F	
TGGGTCCCCCT	CCCTTIGTAT	GGGCGCTCTG	TTTCACTCT	ATTCACTCT	600
G P L	P L Y	G R S V	F T L	F H S	
W V P S	L C M	G A L	F S L Y	F T L	
G S P	P F V W	A L C	F H S	I S L Y	
ATTAAATCTT	GCAACTGAAA	AAAAAAAAAA	AAAAA		635
I K S C	N . K	K K K	K		
L N L	A T E K	K K K	K		
. I L	Q L K	K K K	K		

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
ATGGCCCTCC	CTTATCATA	TTTCTCTTT	ACTGTTCTCT	TACCCCTTT	50
M A L P	Y H T	F L F	T V L L	P P F	
W P S	L I I L	F S L	L F S	Y P L S	
G P P	L S Y	F S L Y	C S L	T P F	
CGCTCTCACT	GCACCCCTC	CATGCTGCTG	TACAACCAGT	AGCTCCCTT	100
A L T	A P P P	C C C	T T S	S S P Y	
L S L	H P L	H A A V	Q P V	A P L	
R S H C	T P S	M L L	Y N Q	L P L	
ACCAAGAGTT	TCTATGAAGA	ACGCGGCTTC	CTGGAAATAT	TGATGCCCCA	150
Q E F	L . R	T R L P	G N I	D A P	
T K S F	Y E E	R G F	L E I L	M P H	
P R V	S M K N	A A S	W K Y	. C P I	
TCATATAGGA	GTTTATCTAA	GGGAAACTCC	ACCTTCACTG	CCCACACCCA	200
S Y R S	L S K	G N S	T F T A	H T H	
H I G	V Y L R	E T P	P S L	P T P I	
I . E	F I .	G K L H	L H C	P H P	
TATGCCCCGC	AACGTCTATA	ACTCTGCCAC	TCTTTGCATG	CATGCAAATA	250
M P R	N C Y N	S A T	L C M	H A N T	
C P A	T A I	T L P L	F A C	M Q I	
Y A P Q	L L .	L C H	S L H A	C K Y	
CTCATTATTG	GACAGGGAAA	ATGATTAAATC	CTAGTTGTCC	TGGAGGACTT	300
H Y W	T G K	M I N P	S C P	G G L	
L I I G	Q G K	. L I	L V V L	E D L	
S L L	D R E N	D . S	. L S	W R T W	
GGAGCCACTG	TCTGTGTGAC	TTACTTCACC	CATACCAGTA	TGCTTGATGG	350
G A T V	C W T	Y F T	H T S M	S D G	
E P L	S V G L	T S P	I P V	C L M G	
S H C	L L D	L L H P	Y Q Y	V . W	

6 / 32
FIG 4 (continued)

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
ACCTCACCTG	TGTAAAATTT	AGCAATACTA	TAGACACAAC	CAGCTCCCAA	750
L T C V K F	S N T I	D T T	S S Q		
T S P V	. N L A I L	. T Q P	A P N		
P H L C K I	. Q Y Y	R H N	Q L P M		
TGCATCAGGT	GGGTACACC	TOCCACACGA	ATAGTCTGCC	TACCTCAGG	800
C I R W	V T P	P T R	I V C L	P S G	
A S G G	. H L P H E	. S A	Y P Q E		
H Q V	G N T	S H T N	S L P	T L R	
AATATTTTTT	GICTGIGGIA	CCTCAGCCTA	TCATTGTTTG	AATGGCTCTT	850
I F F	V C G T	S A Y	H C L	N G S S	
Y F L	S V V	P Q P I	I V	. M A L	
N I F C	L W Y	L S L	S L F E	W L F	
CAGAATCTAT	GIGCTTCCTC	TCATTCTTAG	TGCCCCCTAT	GACCATCTAC	900
E S M	C F L	S F L V	P P M	T I Y	
Q N L C	A S S	H S	. C P L	. P S T	
R I Y	V L P L	I L S	A P Y	D H L H	
ACTGAACAAG	ATTATACAA	TCATGTGGIA	OCTAAGCCCC	ACAACAAAAG	950
T E Q D	L Y N	H V V	P K P H	N K R	
L N K	I Y T I	M S Y	L S P	T T K E	
. T R	F I Q	S C R T	. A P	Q Q K	
AGTACCCATT	CTTCCTTTTG	TTATCAGAGC	AGGAGTGTCTA	GGCAGACTAG	1000
V P I	L P F V	I R A	G V L	G R L G	
Y P F	F L L	L S E Q	E C	. A D	
S T H S	S F C	Y Q S	R S A R	Q T R	
GIACCTGGCAT	TGGCAGTATC	ACAACCTCTA	CTCAGTTCTA	CTACAAACTA	1050
T G I	G S I	T T S T	Q F Y	Y K L	
V L A L	A V S	Q P L	L S S T	T N Y	
Y W H	W Q Y H	N L Y	S V L	L Q T I	

FIG 4 (continued)

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
TCTCAAGAAA	TAAATGGTGA	CATGGAACAG	GTCACCTGACT	CCCTGGTCAC	1100
S Q E I	N G D M E Q	V T D S	L V T		
L K K	. M V T W N R	S L T P W S P			
S R N K W	. H G T G H	. L P G H			
CTTGCAAGAT	CAACTTAACT	CCCTAGCAGC	AGTAGTCTTT	CAAAATCGAA	1150
L Q D Q L N S	L A A V V L	Q N R R			
C K I N L T P	. Q Q . S F K I E				
L A R S T	. L P S S S S P S	K S K			
GAGCTTTAGA	CTTGCTAACC	GCCAAAAGAG	GGGGAACCTG	TTTATTTTTTA	1200
A L D L L T A K R G	G T C L F L				
E L . T C . P P K E	G E P V Y F .				
S F R L A N R	Q K R G N L F I F R				
GGAGAAGAAC	GCTGTATTATTA	TGTTAATCAA	TCCAGAATTG	TCACTGAGAA	1250
G E E R C Y Y V N Q	S R I V T E K				
E K N A V I M L I N	P E L S L R K				
R R T L L L C . S I	Q N C H . E				
AGTTAAAGAA	ATTGAGATC	GAATACAATG	TAGAGCAGAG	GAGCTTCAAA	1300
V K E I R D R I Q C	R A E E L Q N				
L K K F E I E Y N V	E Q R S F K				
S . R N S R S N T M	. S R G A S K				
ACACCGAACG	CTGGGGGCTC	CTCAGCCAAT	GGATGCCCTG	GGTCTCTCCC	1350
T E R W G L L S Q W	M P W V L P				
T P N A G A S S A N	G C P G F S P				
H R T L G P P Q P M	D A L G S P L				
TTCTTAGGAC	CTCTAGCAGC	TCTAATATTG	TTACTCTCTT	TTGGACCTG	1400
F L G P L A A L I L	L L L F G P C				
S . D L . Q L . Y C	Y S S L D P V				
L R T S S S S N I V	T P L W T L				

8 / 32
FIG 4 (continued)

10	20	30	40	50
1234567890	1234567890	1234567890	1234567890	1234567890
TATCTTTAAC	CTCCTTGTA	AGTTTGCTC	TTCAGAATT	GAAGCTGTAA
I F N	L L V K	F V S	S R I	E A V K
S L T	S L L	S L S L	P E L	K L
Y L	P P C	V C L	F Q N	S C K
AGCTACAGAT	GGTCTTACAA	ATGGAACCCC	A	
L Q M	V L Q	M E P		
S Y R W	S Y K	W N P		
A T D	G L T N	G T P		

1450

1481

9 / 32
FIG 5

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
TCAAAATCGA	AGAGCTTTAG	ACTTGCTAAC	CGCCAAAGA	GGGGGAACCT	50
S K S K	S F R	L A N	R Q K R	G N L	
Q N R	R A L D	L L T	A K R	G G T C	
K I E	E L .	T C .	P P K E	G E P	
GTTTATTTTT	AGGGGAAGAA	TGCTGTAGT	ATGTTAATCA	ATCTGGAATC	100
F I F	R G R M	L L V C .	S I W N H		
L F L	G E E C C .	Y V N Q	S G I		
V Y F .	G K N	A V S	M L I N	L E S	
ATTACTGAGA	AAGTTAAAGA	AATTTGAGAT	CGAATATAAT	GTAGAGCAGA	150
Y . E S .	R N L R S	N I M .	S R		
I T E K	V K E I .	D R I .	C R A E		
L L R	K L K K	F E I	E Y N	V E Q R	
GGACCTTCAA	AACACTGCAC	CCTGGGGCCT	CCTCAGCCAA	TGGATGCOCT	200
G P S K	H C T	L G P	P Q P M	D A L	
D L Q	N T A P	W G L	L S Q	W M P W	
T F K	T L H	P G A S	S A N	G C P	
GGACTCTCCC	CTTCTTAGGA	CCTCTAGCAG	CTATAATATT	TTTACTCCTC	250
D S P	L L R T	S S S	Y N I	F T P L	
T L P	F L G	P L A A	I I F	L L L	
G L S P	S . D	L . Q	L . Y F	Y S S	
TTTGGACCOCT	GTATCTTCAA	CTTCCTTGTT	AAGTTTGICT	CTTCAGAAT	300
W T L	Y L Q	L P C .	V C L	F Q N	
F G P C	I F N	F L V	K F V S	S R I	
L D P	V S S T	S L L	S L S	L P E L	
TGAAGCTGTA	AAGCTACAAA	TAGTTCTTCA	AATGGAACCC	CAGATGCAGT	350
. S C K	A T N	S S S	N G T P	D A V	
E A V	K L Q I	V L Q	M E P	Q M Q S	
K L .	S Y K .	F F K	W N P	R C S	

10 / 32
FIG 5 (continued)

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
CCATGACTAA	AATCTACCGT	GGACCCCTGG	ACCGGCCTGC	TAGACTATGC	400
H D .	N L P W	T P G	P A C .	T M L	
M T K	I Y R	G P L D	R P A	R L C	
P . L K	S T V	D P W	T G L L	D Y A	
TCTGATGTTA	ATGACATTGA	AGTCACCCCT	CCCGAGGAAA	TCTCAACTGC	450
. C .	. H .	S H P S	R G N	L N C	
S D V N	D I E	V T P	P E E I	S T A	
L M L	M T L K	S P L	P R K	S Q L H	
ACAACCCCTA	CTACACTCCA	ATTCAGTAGG	AAGCAGTTAG	AGCAGTTGTC	500
T T P T	T L Q	F S R	K Q L E	Q L S	
Q P L	L H S N	S V G	S S .	S S C Q	
N P Y	Y T P	I Q .	E A V R	A V V	
AGCCAACCTC	CCCAACAGTA	CTTGGGTTTT	CCGTGTGAGA	GGGIGGACTG	550
A N L	P N S T	W V F	L L R	G W T E	
P T S	P T V	L G F S	C . E	G G L	
S Q P P	Q Q Y	L G F	P V E R	V D .	
AGAGACAGGA	CTAGCTGGAT	TTCCTAGGCT	GACTAAGAAT	CCCAAGCCT	600
R Q D .	L D F L G .	L R I	P K P		
R D R T	S W I S .	A D .	E S X S L		
E T G	L A G F	P R L	T K N	P X A X	
ANCTGGGAAG	GTGACCGCAT	CCATCTTTAA	ACATGGGGCT	TGCAACTTAG	650
X W E G	D R I H L .	T W G L	Q L S		
X G K	V T A S	I F K	H G A	C N L A	
L G R .	P H	P S L N	M G L	A T .	
CTCACACCCG	ACCAATCAGA	GAGCTCACTA	AAATGCTAAT	CAGGCAAAAA	700
S H P	T N Q R	A H .	N A N	Q A K T	
H T R	P I R	E L T K	M L I	R Q K	
L T P D	Q S E	S S L	K C .	S G K N	

FIG 5 (continued)

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
CAGGAGGTAA	AGCAATAGCC	AATCATCTAT	TGCTGAGAG	CACAGCGGA	750
G G K	A I A	N H L L	P E S	T A G	
Q E V K	Q . P	I I Y	C L R A	Q R E	
R R .	S N S Q	S S I	A . E	H S G K	
AGGACAAGGA	TTGGGATATA	AATCAGGCA	TTCAGCCAG	CAACAGCAAC	800
R T R I	G I .	T Q A	F K P A	T A T	
G Q G	L G Y K	L R H	S S Q	Q Q Q P	
D K D	W D I	N S G I	Q A S	N S N	
CCCCCTTGGG	TCCCCGCCA	TTGTATGGGA	GCTCTGTTTT	CACCTATTT	850
P F G	S P P I	V W E	L C F	H S I S	
P L G	P L P	L Y G S	S V F	T L F	
P L W V	P S H	C M G	A L F S	L Y F	
CACCTATTA	AATCATGCAA	CTGCACCTT	CTGGTCGGTG	TTTTTATGG	900
L Y .	I M Q	L H S S	G P C	F L W	
H S I K	S C N	C T L	L V R V	F Y G	
T L L	N H A T	A L F	W S V	F F M A	
CTCAAGCTGA	GCTTTTGTC	GCCATCCACC	ACTGCTGTTT	GCCACCGTCA	950
L K L S	F C S	P S T	T A V C	H R H	
S S .	A F V R	H P P	L L F	A T V T	
Q A E	L L F	A I H H	C C L	P P S	
CAGACCGCT	GCTGACTTCC	ATCCCTTGG	ATCCAGCAGA	GTTGTCCTG	1000
R P A	A D F H	P F G	S S R	V S T V	
D P L	L T S	I P L D	P A E	C P L	
Q T R C	. L P	S L W	I Q Q S	V H C	
TGCTCTGAT	CCAGCGAGGT	ACCCATTGCC	ACTCCCGATC	AGGCTAAAGG	1050
L L I	Q R G	T H C H	S R S	G . R	
C S .	S S E V	P I A	T P D Q	A K G	
A P D	P A R Y	P L P	L P I	R L K A	

12 / 32
FIG 5 (continued)

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
CTTGCCATTG	TTCTGTCATG	GCTAAGTGCC	TGGGTTTGTC	CTAATAGAAC	1100
L A I V	P A W	L S A	W V C P	N R T	
L P L	F L H G	. V P	G F V	L I E L	
C H C	S C M	A K C L	G L S	. . N	
TGAACACTGG	TCACTGGGTT	CCATGGTTCT	CTTCCATGAC	CCACGGCTTC	1150
E H W	S L G S	M V L	F H D	P R L L	
N T G	H W V	P W F S	S M T	H G F	
. T L V	T G F	H G S	L P .	P T A S	
TAATAGAGCT	ATAACACTCA	CCGCATGGCC	CAAGATTCCA	TTCTTTGGTA	1200
I E L	. H S	P H G P	R F H	S L V	
. . S Y	N T H	R M A	Q D S I	P W Y	
N R A	I T L T	A W P	K I P	F L G I	
TCTGTGAGGC	CAAGAACCCC	AGGTCAGAGA	ANGTGAGGCT	TGCCACCATT	1250
S V R P	R T P	G Q. R	X . G L	P P F	
L . G	Q E P Q	V R E	X E A	C H H L	
C E A	K N P	R S E X	V R L	A T I	
TGGGAAGTGG	CCCACTGCCA	TTTGTGGTAGC	GGCCCACCAC	CATCTTGGGA	1300
G K W	P T A I	L V A	A H H	H L G S	
G S G	P L P	F W .	R P T T	I L G	
W E V A	H C H	F G S	G P P P	S W E	
GCTGTGGGAG	CAAGGATCCC	CCAGTAACA			1329
C G S	K D P	P V T			
A V G A	R I P	Q .			
L W E	Q G S P	S N			

13 / 32

FIG 6

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
CCTAGAACGT	ATTCTGGAGA	ATTGGGACCA	ATGTGACACT	CAGACGCTAA	50
P R T Y	S G E	L G P	M . H S	D A K	
L E R	I L E N	W D Q	C D T	Q T L R	
. N V	F W R	I G T N	V T L	R R .	
GAAAGAAAG	ATTTATATTC	TTCTGCAGTA	CCGCTGGGC	ACAATATCCT	100
K E T	I Y I L	L Q Y	R L A	T I S S	
K K R	F I F	F C S T	A W P	Q Y P	
E R N D	L Y S	S A V	P P G H	N I L	
CTTCAAGGA	GAGAAACCTG	GCTTCTGAG	GGAAGTATAA	ATTATAACAT	150
S R E	R N L	A S . G	K Y K	L . H	
L Q G R	E T W	L P E	G S I N	Y N I	
F K G	E K P G	F L R	E V .	I I T S	
CATCTTACAG	CTAGACCTCT	TCTGTAGAAA	GGAGGGCAAA	TGGAGTGAAG	200
H L T A	R P L	L . K	G G Q M	E . S	
I L Q	L D L F	C R K	E G K	W S E V	
S Y S	. T S	S V E R	R A N	G V K	
TGCCATATGT	GCAAACCTTC	TTTTCATTA	GAGACAATC	ACAATTATGT	250
A I C	A N F L	F I K	R Q L	T I M .	
P Y V	Q T F	F S L R	D N S	Q L C	
C H M C	K L S	F H .	E T T H	N Y V	
AAAAAGTGTG	GTTTATGCCC	TACAGGAAGC	CCTCAGAGTC	CACTTCCCTA	300
K V W	F M P	Y R K P	S E S	T S L	
K K C G	L C P	T G S	P Q S P	P P Y	
K S V	V Y A L	Q E A	L R V	H L P T	
CCCCAGGTC	CCCTCCCCGA	CTCCTTCTC	AACATAAAG	GACCCCCCTT	350
P Q R P	L P D	S F L	N . . G	P P F	
P S V	P S P T	P S S	T N K	D P P L	
P A S	P P R	L L P Q	L I R	T P L	
TATCCCAAC	GGTCCAAAAG	GAGATAGACA	AAGGGGTAAA	CAATGAACCA	400
N P N	G P K G	D R Q	R G K	Q . T K	
T Q T	V Q K	E I D K	G V N	N E P	
. P K R	S K R	R . T	K G .	T M N Q	
AAGAGTGCCA	ATATTCCCCG	ATTATGCCCC	CTCCAAGCAG	TGAGAGGAGG	450
E C Q	Y S P	I M P P	P S S	E R R	
K S A N	I P R	L C P	L Q A V	R G G	
R V P	I F P D	Y A P	S K Q .	E E E	
AGAATTGGC	CCAGCCAGAG	TGCTGTACC	TTTTTCTCTC	TCAGACTTAA	500
R I R P	S Q S	A C T	F F S L	R L K	
E F G	P A R V	P V P	F S L	S D L K	
N S A	Q P E	C L Y L	F L S	Q T .	

14 / 32
FIG 6 (continued)

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
AGCAAATTAA	AATAGACCTA	GGTAAATTCT	CAGATAACCC	TGACGGCTAT	550
A N .	N R P R .	I L R .	P .	R L Y	
Q I K	I D L	G K F S	D N P	D G Y	
S K L K .	T .	V N S	Q I T L	T A I	
ATTGATGTTT	TACAAGGGTT	AGGACAATCC	TTTGATCTGA	CATGGAGAGA	600
. C F	T R V	R T I L .	S D	M E R	
I D V L	Q G L	G Q S	F D L T	W R D	
L M F	Y K G .	D N P	L I .	H G E I	
TATAATGITA	CTACTAAATC	AGACACTAAC	CCCAAATGAG	AGAAGTGGCG	650
Y N V T	T K S	D T N	P K .	E K C R	
I M L	L L N Q .	T L T	P N E	R S A A	
. C Y Y .	I R H .	P	Q M R	E V P	
CTGTAACTGC	AGCCCGAGAG	TTTGGCGATC	TTTGGTATCT	CAGTCAGGCC	700
C N C	S P R V	W R S	L V S	Q S G Q	
V T A	A R E	F G D L	W Y L	S Q A	
L .	L Q	P E S	L A I	F G I S .	V R P
AACAATAGGA	TGACAACAGA	GGAAAGAACA	ACTGCCACAG	GCCAGCAGGC	750
Q .	D D N R	G K N N	S H R	P A G	
N N R M	T T E	E R T	T P T G	Q Q A	
T I G .	Q Q R	K E Q	L P Q	A S R Q	
AGTTCCCACT	GTAGACCTTC	ATTGGGACAC	AGAATCAGAA	CATGGAGATT	800
S S Q C	R P S	L G H	R I R T	W R L	
V P S	V D P H	W D T	E S E	H G D W	
F P V .	T L	I G T Q	N Q N	M E I	
GGTGGCACAA	ACATTTGCTA	ACTTGGGTGC	TAGAAGGACT	GAGGAAACT	850
V P Q	T F A N	L R A	R R T	E E N .	
C H K	H L L	T C V L	E G L	R K T	
G A T N	I C .	L A C .	K D .	G K L	
AGGAGAAGC	CTATGAATTA	CTCAATGATG	TOCACTATAA	CACAGGGAAA	900
E E A	Y E L	L N D V	H Y N	T G K	
R K K P	M N Y	S M M	S T I T	Q G K	
G R S	L .	I T Q .	C P L .	H R E R	
GGAGAAAAT	CTTACTGCTT	TTCCTGGACAG	ACTAAGGGAG	GCATTGAGGA	950
G R K S	Y C F	S G Q	T K G G	I E E	
E E N	L T A F	L D R	L R E	A L R K	
K K I	L L L	F W T D .	G R H .	G	
AGCATACCTC	CCTGTCACTT	GACTCTATTG	AAGGCCAACT	AATCTTAAAG	1000
A Y L	P V T .	L Y .	R P T	N L K G	
H T S	L S P	D S I E	G Q L	I L K	
S I P P	C H L	T L L	K A N .	S . R	

15 / 32
FIG 6 (continued)

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
GATAAGTTTA	TCACTCAGTC	AGCTGCAGAC	ATTAGAAAAA	ACTTCAAAAG	1050
. V Y H S V	S C R H	. K K	L Q K		
D K F I	T Q S A A D	I R K N	F K S		
I S L S L S Q	L Q T	L E K	T S K V		
TCTGCTTAG	GCCCCGAGCA	GAACTTAGAA	ACCTTATTTA	ACTTGGCATC	1100
S A L G	P E Q N L E	T L F N	L A S		
L P .	A R S R T .	K P Y L	T W H P		
C L R	P G A E L R N	P I .	L G I		
CTCAGTTTTT	TATAATAGAG	ATCAGGAGGA	GCAGGGGAAA	CGGGACAAAC	1150
S V F	Y N R D	Q E E	Q A K	R D K R	
Q F F	I I E	I R R S	R R N	G T N	
L S F L	. . R	S G G	A G E T	G Q T	
GGGATAAAAA	AAAAAGGGGG	GGTCCACTAC	TTTAGTCATG	GGCCTCAGGC	1200
D K K	K R G	G P L L	. S W	P S G	
G I K K	K G G	V H Y	F S H	G P Q A	
G .	K K K G G	S T T	L V M	A L R Q	
AAGCAGACTT	TGGAGGCTCT	GCAAAAGGGA	AAAGCTGGGC	AAATCAAATG	1250
K Q T L	E A L	Q K G	K A G Q	I K C	
S R L	W R L C	K R E	K L G	K S N A	
A D F	G G S	A K G K	S W A	N Q M	
CCTAATAGGG	CTGGCTTCCA	GTGGGGTCTA	CAAGGACACT	TTAAAAAGA	1300
L I G	L A S S	A V Y	K D T	L K K I	
. .	G W L P	V R S T	R T L	. K R	
P N R A	G F Q	C G L	Q G H F	K K D	
TTATCCAAGT	AGAAATAAGC	CGCCCCCTTG	TCCATGCCCC	TTAGGTCAAG	1350
I Q V	E I S	R P L V	H A P	Y V K	
L S K .	K . A	A P L	S M P L	T S R	
Y P S	R N K P	P P C	P C P	L R Q G	
GGATCACTG	GAAGGCCCCAC	TGCCCCAGGG	GATGAAGATA	CTCTGAGTCA	1400
G I T G	R P T A P G	D E D T	L S Q		
E S L	E G P L	P Q G	M K I	L . V R	
N H W	K A H	C P R G	. R Y	S E S	
GAAGCCATT	ACCAGATGAT	CCAGCAGCAG	GACTGAGGGT	GGGGGGGGGG	1450
K P L	T R .	S S S R	T E G	A R G E	
S H .	P D D	P A A G	L R V	P G A	
E A I N	Q M I	Q Q Q	D .	G C P G R	
AGGCGCAGGC	CATGCCATCA	CCCTCAGAGA	GGGGGGGGTA	TGTTTGAACA	1500
R Q P	M P S	P S Q S	P G Y	V . P	
S A S P	C H H	P H R	A P G M	F D H	
A P A	H A I T	L T E	P R V	C L T I	

16 / 32

FIG 6 (continued)

10	20	30	40	50
1234567890	1234567890	1234567890	1234567890	1234567890
TTGAGAGCCA A				
				1511
L R A				
. E P				
E S Q				

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
ATGGGCAGCA	GCCATCATCA	TCATCATCAC	AGCAGCGGCC	TGGTGGCGCG	50
M G S S	H H H	H H H	S S G L	V P R	
CGGCAGCCAT	ATGGCTAGCA	TGACTGGTGG	ACAGCAAATG	GGTGGGATCC	100
G S H	M A S M	T G G	Q Q M	G R I L	
TAGAAGTAT	TCTGGAGAAT	TGGGACCAAT	GTGACACTCA	GACGCTAAGA	150
E R I	L E N	W D Q C	D T Q	T L R	
AAGAAAGAT	TTATATTCTT	CTGCAGTACC	GCGTGGCCAC	AATATCTCT	200
K K R F	I F F	C S T	A W P Q	Y P L	
TCAAGGGAGA	GAAACCTGGC	TTCCTGAGGG	AAGTATAAAT	TATAACATCA	250
Q G R	E T W L	P E G	S I N	Y N I I	
TCTTACAGCT	AGACCTCTTC	TGTAGAAAGG	AGGGCAAATG	GAGTGAAGTG	300
L Q L	D L F	C R K E	G K W	S E V	
CCATATGTGC	AACTTTCTT	TTCATTAAGA	GACAACTCAC	AATTATGTAA	350
P Y V Q	T F F	S L R	D N S Q	L C K	
AAAGTGTGGT	TTATGCCCTA	CAGGAAGCCC	TCAGAGTCCA	CTTCCCTACC	400
K C G	L C P T	G S P	Q S P	P P Y P	
CCAGGCTCC	CTCCCCGACT	CCTTCTCTCA	CTAATAAGGA	CCCCCTTTA	450
S V P	S P T	P S S T	N K D	P P L	
ACCCAAAGG	TCCAAAAGGA	GATAGACAAA	GGGGTAAACA	ATGAACCAAA	500
T Q T V	Q K E	I D K	G V N N	E P K	
GAGTGCCAAT	ATTCCCCGAT	TATGCCCCCT	CCAAGCAGTG	AGAGGAGGAG	550
S A N	I P R L	C P L	Q A V	R G G E	
AATTCGGCC	AGCCAGAGTG	CCTGTACCTT	TTTCTCTCTC	AGACTTAAAG	600
F G P	A R V	P V P F	S L S	D L K	
CAAATTAAAA	TAGACCTAGG	TAAATTCTCA	GATAACCTTG	ACGGCTATAT	650
Q I K I	D L G	K F S	D N P D	G Y I	
TGATGTTTAA	CAAGGGTTAG	GACAATCTTT	TGATCTGACA	TGGAGAGATA	700
D V L	Q G L G	Q S F	D L T	W R D I	
TAATGTTACT	ACTAAATCAG	ACACTAACC	CAAATGAGAG	AAGTGGCGCT	750
M L L	L N Q	T L T P	N E R	S A A	
GTAACTGCAG	CCCGAGAGTT	TGGCGATCTT	TGGTATCTCA	GTGAGGCCAA	800
V T A A	R E F	G D L	W Y L S	Q A N	

18 / 32
FIG 7 (continued)

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
CAATAGGATG	ACAACAGAGG	AAAGAACAAC	TCCACAGGC	CAGCAGGCAG	850
N R M	T T E E	R T T	P T G	Q Q A V	
TTCCAGTGT	AGACCTCAT	TGGGACACAG	AATCAGAACA	TGGAGATTGG	900
P S V	D P H	W D T E	S E H	G D W	
TGCCACAAAC	ATTTGCTAAC	TTGGTGCTA	GAAGGACTGA	GGAAACTAG	950
C H K H	L L T	C V L	E G L R	K T R	
GAAGAAGCCT	ATGAATTACT	CAATGATGTC	CACTATAACA	CAGGGAAAGG	1000
K K P	M N Y S	M M S	T I T	Q G K E	
AAGAAAATCT	TACTGCTTTT	CTGGACAGAC	TAAGGGAGGC	ATTGAGGAAG	1050
E N L	T A F	L D R L	R E A	L R K	
CATACCTCCC	TGTCACCTGA	CTCTATTGAA	GGCCAACTAA	TCTTAAAGGA	1100
H T S L	S P D	S I E	G Q L I	L K D	
TAAGTTTATC	ACTCAGTCAG	CTGCAGACAT	TAGAAAAAAC	TTCAAAAGTC	1150
K F I	T Q S A	A D I	R K N	F K S L	
TGCCTAAGCT	TGGGGGCGCA	CTCGAGCACC	ACCACCACCA	CCACTGAGAT	1200
P K L	A A A	L E H H	H H H	H . D	
CCGGCTGCTA	ACAAAGCCCG	AAAGGAAGCT	GAGTTGGCTIN	GTGGCNA	1247
P A A N	K A R	K E A	E L A X	G	

19 / 32
FIG 8

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
ATGGCTAGCA	TGACTGGTGG	ACAGCAAATG	GGTCGGATCC	TAGAACGTAT	50
M A S M	T G G	Q Q M	G R I L	E R I	
TCTGGAGAAT	TGGGACCAAT	GTGACACTCA	GACGCTAAGA	AAGAAACGAT	100
L E N	W D Q C	D T Q	T L R	K K R F	
TTATATTCTT	CTGCAGTACC	GOCTGGCCAC	AATATCTCTT	TCAAGGGAGA	150
I F F	C S T	A W P Q	Y P L	Q G R	
GAAACCTGGC	TTCTTGAGGG	AAGTATAAAT	TATAACATCA	TCTTACAGCT	200
E T W L	P E G	S I N	Y N I I	L Q L	
AGACCTCTTC	TGTAGAAAGG	AGGGCAAATG	GAGTGAAGTG	CCATATGTGC	250
D L F	C R K E	G K W	S E V	P Y V Q	
AAACTTTCTT	TTTATTAAGA	GACAATCAC	AATTATGTAA	AAAGTGTGGT	300
T F F	S L R	D N S Q	L C K	K C G	
TTATGCCCCA	CAGGAAGCCC	TCAGAGTCCA	CTCTCTTACC	CCAGCGTCCC	350
L C P T	G S P	Q S P	P P Y P	S V P	
CTCCCCGACT	CTTCTCTCAA	CTAATAAGGA	CCCCCTTTA	ACCCAAACGG	400
S P T	P S S T	N K D	P P L	T Q T V	
TCCAAAAGGA	GATAGACAAA	GGGGTAAACA	ATGAACCAAA	GAGTGCCAAT	450
Q K E	I D K	G V N N	E P K	S A N	
ATTCCCCGAT	TATGCCCCCT	CCAAGCAGTG	AGAGGAGGAG	AATTGGGCCC	500
I P R L	C P L	Q A V	R G G E	F G P	
AGCCAGAGTG	CTGTACCTT	TTTCTCTCTC	AGACTTAAAG	CAAATTAAAA	550
A R V	P V P F	S L S	D L K	Q I K I	
TAGACTAGG	TAAATTCTCA	GATAACCTTG	ACGGCTATAT	TGATGTTTAA	600
D L G	K F S	D N P D	G Y I	D V L	
CAAGGGTTAG	GACAATCTTT	TGATCTGACA	TGGAGAGATA	TAATGTTACT	650
Q G L G	Q S F	D L T	W R D I	M L L	
ACTAAATCAG	AACTAAACCC	CAAATGAGAG	AAGTGGCGCT	GTAATGTCAG	700
L N Q	T L T P	N E R	S A A	V T A A	
CCCGAGAGTT	TGGGATCTTT	TGGTATCTCA	GTGAGGOCOA	CAATAGGATG	750
R E F	G D L	W Y L S	Q A N	N R M	
ACAACAGAGG	AAAGAACAAC	TCCACAGGC	CAGCAGGCAG	TTCCAGTGT	800
T T E E	R T T	P T G	Q Q A V	P S V	

20 / 32

FIG 8 (continued)

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
AGACCCATCAT	TGGGACACAG	AATCAGAACA	TGGAGATTGG	TGCCACAAAC	850
D P H	W D T E	S E H	G D W	C H K H	
ATTTGCTAAC	TTGGGTGCTA	GAAGGACTGA	GGAAACTAG	GAAGAAGCCT	900
L L T	C V L	E G L R	K T R	K K P	
ATGAATTACT	CAATGATGTC	CACTATAACA	CAGGGAAAGG	AAGAAAATCT	950
M N Y S	M M S	T I T	Q G K E	E N L	
TACTGCTTTT	CTGGACAGAC	TAAGGGAGGC	ATTGAGGAAG	CATACCTCCC	1000
T A F	L D R L	R E A	L R K	H T S L	
TGTCACCTGA	CTCTATTGAA	GGCCAACTAA	TCTTAAAGGA	TAAGTTTATC	1050
S P D	S I E	G Q L I	L K D	K F I	
ACTCAGTCAG	CTGCAGACAT	TAGAAAAAAC	TTCAAAAGTC	TGCCTAAGCT	1100
T Q S A	A D I	R K N	F K S L	P K L	
TGGGGGCGCA	CTGGAGCACC	ACCAACCAACA	CCACTGAGAT	CCGGCTGCTA	1150
A A A	L E H H	H H H	H . D	P A A N	
ACAAAGCCCCG	AAAGGAAGCT	GAGTTGGCTG	GTGGCA		1186
K A R	K E A	E L A G	G		

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
TGTCCTGCTGT	GCTCCTGATC	CAGCACAGGC	GCCCATTTGCC	TCTCCCAATT	50
C P L C S . S	S T G A H C L	S Q L			
V R C A P D P	A Q A P I A	S P N W			
S A V L L I	Q H R R	P L P L P I			
GGGCTAAAGG	CTTGCCATTG	TTCCTGCACA	GCTAAGTGGC	TGGGTTTCATC	100
G . R L A I V	P A Q L S A	W V H P			
A K G L P L	F L H S . V P	G F I			
G L K A C H C	S C T A K C L	G S S			
CTAATCGAGC	TGAACACTAG	TCACTGGGTT	CCACGGTTCT	CTTCCATGAC	150
N R A E H .	S L G S T V L	F H D			
L I E L N T S	H W V P R F S	S M T			
. S S . T L V	T G F H G S	L P . P			
CCATGGCTTC	TAATAGAGCT	ATAACACTCA	CIGCATGGTC	CAAGATTCCA	200
P W L L I E L .	H S L H G P	R F H			
H G F . . S Y	N T H C M V	Q D S I			
M A S N R A	I T L T A W S	K I P			
TTCCTTGGA	TCCGTGAGAC	CAAGAACCCC	AGGTCAGAGA	ACACAAGGCT	250
S L E S V R P	R T P G Q R	T Q G L			
P W N P . D	Q E P Q V R E	H K A			
F L G I R E T	K N P R S E N	T R L			
TGCCAOCATG	TTGGAAGCAG	CCCACCACCA	TTTGTGAAGC	AGCCCGCCAC	300
P P C W K Q	P T T I L E A	A R H			
C H H V G S S	P P P F W K Q	P A T			
A T M L E A A	H H H F G S	S P P L			
TATCTTGGGA	GCTCTGGGAG	CAAGGACCCC	AGGTAACAAT	TTGGTGACCA	350
Y L G S S G S	K D P R . Q F	G D H			
I L G A L G A	R T P G N N	L V T T			
S W E L W E	Q G P Q V T I	W . P			
CGAAGGGACC	TGAATCCGCA	ACCATGAAGG	GATCTCCAAA	GCAATTGGAA	400
E G T . I R N	H E G I S K	A I G N			
K G P E S A	T M K G S P K	Q L E			
R R D L N P Q	P . R D L	Q S N W K			

22 / 32
FIG 9 (continued)

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
ATGTTCTCTCC	CAAGGCAAAA	ATGCCCCCTAA	GATGTATTCT	GGAGAATTGG	450
V P P	K A K	M P L R	C I L	E N W	
M F L P	R Q K	C P .	D V F W	R I G	
C S S	Q G K N	A P K	M Y S	G E L G	
GACCAATTIG	ACCCTCAGAC	AGTAAGAAAA	AAATGACTTA	TATTCTTCTG	500
D Q F D	P Q T	V R K	K . L I	F F C	
T N L	T L R Q	. E K	N D L	Y S S A	
P I .	P S D	S K K K	M T Y	I L L	
CAGTACCGCC	CTGGCCACGA	TATCCTCTTC	AAGGGGGAGA	AACCTGGCCT	550
S T A	L A T I	S S S	R G R	N L A S	
V P P	W P R	Y P L Q	G G E	T W P	
Q Y R P	G H D	I L F	K G E K	P G L	
CCTGAGGGAA	GTATAAATTA	TAACACCATC	TTACAGCTAG	ACCTGTTTIG	600
. G K	Y K L	. H H L	T A R	P V L	
P E G S	I N Y	N T I	L Q L D	L F C	
L R E V	. I I	T P S	Y S .	T C F V	
TAGAAAAGGA	GGCAAATGGA	GIGAAGIGCC	ATATTIACAA	ACITTCITTT	650
. K R R	Q M E	. S A	I F T N	F L F	
R K G	G K W S	E V P	Y L Q	T F F S	
E K E	A N G	V K C H	I Y K	L S F	
CATTAAAAGA	CAACTCGCAA	TTATGTAAAC	AGIGTGATTT	GIGTTCCTAC	700
I K R	Q L A I	M L T V	. F	V F L H	
L K D	N S Q	L C .	Q C D L	C S Y	
H . K T	T R N	Y V N	S V I C	V P T	
ACGGAAGCCC	TCAGATTCTA	CTCCCCACCC	CCGGCATCTC	CCCTGAATCC	750
G S P	Q I L	L P T P	G I S	P E S	
T E A L	R F Y	S P P	P A S P	L N P	
R K P	S D S T	P H P	R H L	P . I P	
CTCCCCAACT	TATT				764
L P N L					
S P T Y					
P Q L I					

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
TGTCGCTGT	GCTCCTGATC	CAGCACAGGC	GGCCATIGCC	TCTCCCAATT	50
C P L C S . S	S T G A H C L	S Q L			
V R C A P D P	A Q A P I A	S P N W			
S A V L L I	Q H R R	P L P L P I			
GGGCTAAAGG	CTTGCCATTG	TTCCTGCACA	GCTAAGTGCC	TGGGTTCATC	100
G . R L A I V	P A Q L S A	W V H P			
A K G L P L	F L H S . V	P G F I			
G L K A C H C	S C T A K C L	G S S			
CTAATCGAGC	TGAACACTAG	TCACTGGGTT	CCACGGTTCT	CTTCCATGAC	150
N R A E H .	S L G S T V L	F H D			
L I E L N T S	H W V P R F S	S M T			
. S S . T L V	T G F H G S	L P . P			
CCATGGCTTC	TAATAGAGCT	ATAACACTCA	CTGCATGGTC	CAAGATTCCA	200
P W L L I E L .	H S L H G P	R F H			
H G F . . S Y	N T H C M V	Q D S I			
M A S N R A	I T L T A W S	K I P			
TTCCTTGGAA	TCGTGAGAC	CAAGAACCCC	AGGTCAGAGA	ACACAAGGCT	250
S L E S V R P	R T P G Q R	T Q G L			
P W N P . D	Q E P Q V R E	H K A			
F L G I R E T	K N P R S E N	T R L			
TGCCACCATG	TTGGAAGCAG	CCCACCACCA	TTTIGGAAGC	GGCCCCGCAC	300
P P C W K Q	P T T I L E A	A R H			
C H H V G S S	P P P F W K R	P A T			
A T M L E A A	H H H F G S	G P P L			
TATCTTGGGA	GCTCTGGGAG	CAAGGACCCC	CAGGTAACAA	TTTGGTGACC	350
Y L G S S G S	K D P Q V T I	W . P			
I L G A L G A	R T P R . Q	F G D H			
S W E L W E	Q G P P G N N	L V T			
ACGAAGGGAC	CTGAATCGC	AACCATGAAG	GGATCTCCAA	AGCAATTGGA	400
R R D L N P Q	P . R D L Q	S N W K			
E G T . I R	N H E G I S K	A I G			
T K G P E S A	T M K G S P K	Q L E			

24 / 32
FIG 10 (continued)

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
AATGTTCCCTC	CCAAGGCAAA	AATGCCCCCTA	AGATGTATTC	TGGAGAATTG	450
C S S	Q G K	N A P K	M Y S	G E L	
N V P P	K A K	M P L	R C I L	E N W	
M F L	P R Q K	C P .	D V F	W R I G	
GGACCAATCT	GACCCCTCAGA	CAGTAAGAAA	AAAAATGACT	TATATTCTTC	500
G P I .	P S D	S K K	K N D L	Y S S	
D Q S	D P Q T	V R K	K M T	Y I L L	
T N L	T L R	Q .	E K K .	L I F F	
TGCAGTACCG	CCTGGCCACG	GATATCCTCT	TCAAGGGGGA	GAAACCTGGC	550
A V P	P G H G	Y P L	Q G G	E T W P	
Q Y R	L A T	D I L F	K G E	K P G	
C S T A	W P R	I S S	S R G R	N L A	
CTCCTGAGGG	AAGTATAAAT	TATAACACCA	TCTTACAGCT	AGACCTGTTT	600
P E G	S I N	Y N T I	L Q L	D L F	
L L R E	V .	I I T P	S Y S .	T C F	
S .	G K Y K L	. H H	L T A	R P V L	
TGTAGAAAAG	GAGGCAAATG	GAGTGAAGTG	CCATATTTAC	AAACTTTCTT	650
C R K G	G K W	S E V	P Y L Q	T F F	
V E K	E A N G	V K C	H I Y	K L S F	
. K R	R Q M E .	S A	I F T	N F L	
TTCATTAAAA	GACAACTCGC	AATTATGTAA	ACAGTGTGAT	TTGTGTCTTA	700
S L K	D N S Q	L C K	Q C D	L C P T	
H .	K T T R	N Y V N	S V I	C V L	
F I K R	Q L A	I M .	T V .	F V S Y	
CAGGAAGCCC	TCAGATCTAC	CTCCCTACCC	CGGCATCTCC	CTGACTCCTT	750
G S P	Q I Y	L P T P	A S P .	L L	
Q E A L	R S T	S L P	R H L P	D S F	
R K P	S D L P	P Y P	G I S	L T P S	
CCCCAACTAA	TAAGGACCCA	CTTCAGCCCA	AACAGTCCAA	AAGGACATAG	800
P Q L I	R T H	F S P	N S P K	G H	
P N .	. G P T	S A Q	T V Q	K D I	
P T N	K D P	L Q P K	Q S K	R T .	

25 / 32

FIG 11

10	20	30	40	50
1234567890	1234567890	1234567890	1234567890	1234567890
GGCATTGATA	GCACCCATCA	GATGGCCAAA	TCATTATTIA	CTGGACCAGG
G I D S	T H Q	M A K	S L F T	G P G
A L I	A P I R	W P N	H Y L	L D Q A
H . .	H P S	D G Q I	I I Y	W T R
50				
CCTTTTCAAA	ACTATCAAGC	AGATAGGGCC	CGTGAAGCAT	GCCAAAGAAA
L F K	T I K Q	I G P	V K H	A K E I
F S K	L S S	R . G P	. S M	P K K
P F Q N	Y Q A	D R A	R E A C	Q R N
100				
TAATCCCCCTG	CCTTATCGCC	ATGTTCCCTC	AGGAGAACAA	AGAACAGGCC
I P C	L I A	M F L Q	E N K	E Q A
. S P A	L S P	C S F	R R T K	N R P
N P L	P Y R H	V P S	G E Q	R T G H
150				
ATTACCCAGG	GGAAGACTGG	CAACTAGATT	TTACCCACAT	GGCCAAATGT
I T Q G	K T G	N . I	L P T W	P N V
L P R	G R L A	T R F	Y P H	G Q M S
Y P G	E D W	Q L D F	T H M	A K C
200				
CAGGGATTTC	AGCATCTACT	AGTCTGGGCA	GATACTTTCA	CTGGTTGGGT
R D F	S I Y .	S G Q	I L S	L V G W
G I S	A S T	S L G R	Y F H	W L G
Q G F Q	H L L	V W A	D T F T	G W V
250				
GGAGICTTCT	CCTTGTAGGA	CAGAAAAGAC	CCAAGAGGTA	ATAAAGGCAC
S L L	L V G	Q K R P	K R .	. R H
G V F S	L . D	R K D	P R G N	K G T
E S S	P C R T	E K T	Q E V	I K A L
300				
TAATGAAATA	ATTCCAGAT	TTGGACTTCC	CCCAGGATTA	CAGGGTGACA
. . N N	S Q I	W T S	P R I T	G . Q
N E I	I P R F	G L P	P G L	Q G D N
M K .	F P D	L D F P	Q D Y	R V T
350				

26 / 32
FIG 11 (continued)

10	20	30	40	50	
1234567890	1234567890	1234567890	1234567890	1234567890	
ATGGCCCCGC	TTTCAAGGCT	GCAGTAACCC	AGGGAGTATC	CCAGGTGTTA	400
W P R	F Q G C	S N P	G S I	P G V R	
G P A	F K A	A V T Q	G V S	Q V L	
M A P L	S R L	Q .	P R E Y P	R C .	
GGCATACAAT	ATCACTTACA	CTGIGCCTGG	AGGCCACAAT	CCTCCAGAAA	450
H T I	S L T	L C L E	A T I	L Q K	
G I Q Y	H L H	C A W	R P Q S	S R K	
A Y N	I T Y T	V P G	G H N	P P E K	
AGTCAAGAAA	ATGAATGAAA	CACTCAAAGA	TCTAAAAAAG	CTAACCCAAG	500
S Q E N	E .	N T Q R	S K K A	N P R	
V K K	M N E T	L K D	L K K	L T Q E	
S R K	. M K	H S K I	. K S	. P K	
AAACCCACAT	TGCATGACCT	GTTCTGTTGC	CTATAACCTT	ACTAAGAATC	550
N P H	C M T C	S V A	Y N L	T K N P	
T H I	A .	P V L L P	I T L	L R I	
K P T L	H D L	F C C	L .	P Y . E S	
CATAACTATC	CCCCAAAAG	CAGGACTTAG	CCCATAAGAG	ATGCTATATG	600
. L S	P K K	Q D L A	H T R	C Y M	
H N Y P	P K S	R T .	P I R D	A I W	
I T I	P Q K A	G L S	P Y E	M L Y G	
GATGGCCTTT	CCTAACCAAT	GACCTTGTC	TTGACTGAGA	AATGGCCAAC	650
D G L S	. P M	T L C	L T E K	W P T	
M A F	P N Q .	P C A	. L R	N G Q L	
W P F	L T N	D L V L	D .	E M A N	
TTAGTTGCAG	ACATCACCTC	CTTAGCCAAA	TATCAACAAG	TTCTTAAAAC	700
. L Q	T S P P	. P N	I N K	F L K H	
S C R	H H L	L S Q I	S T S	S . N	
L V A D	I T S	L A K	Y Q Q V	L K T	

27 / 32

GTGACATG 758
V T
. H
D M

758

V T

. H

D M

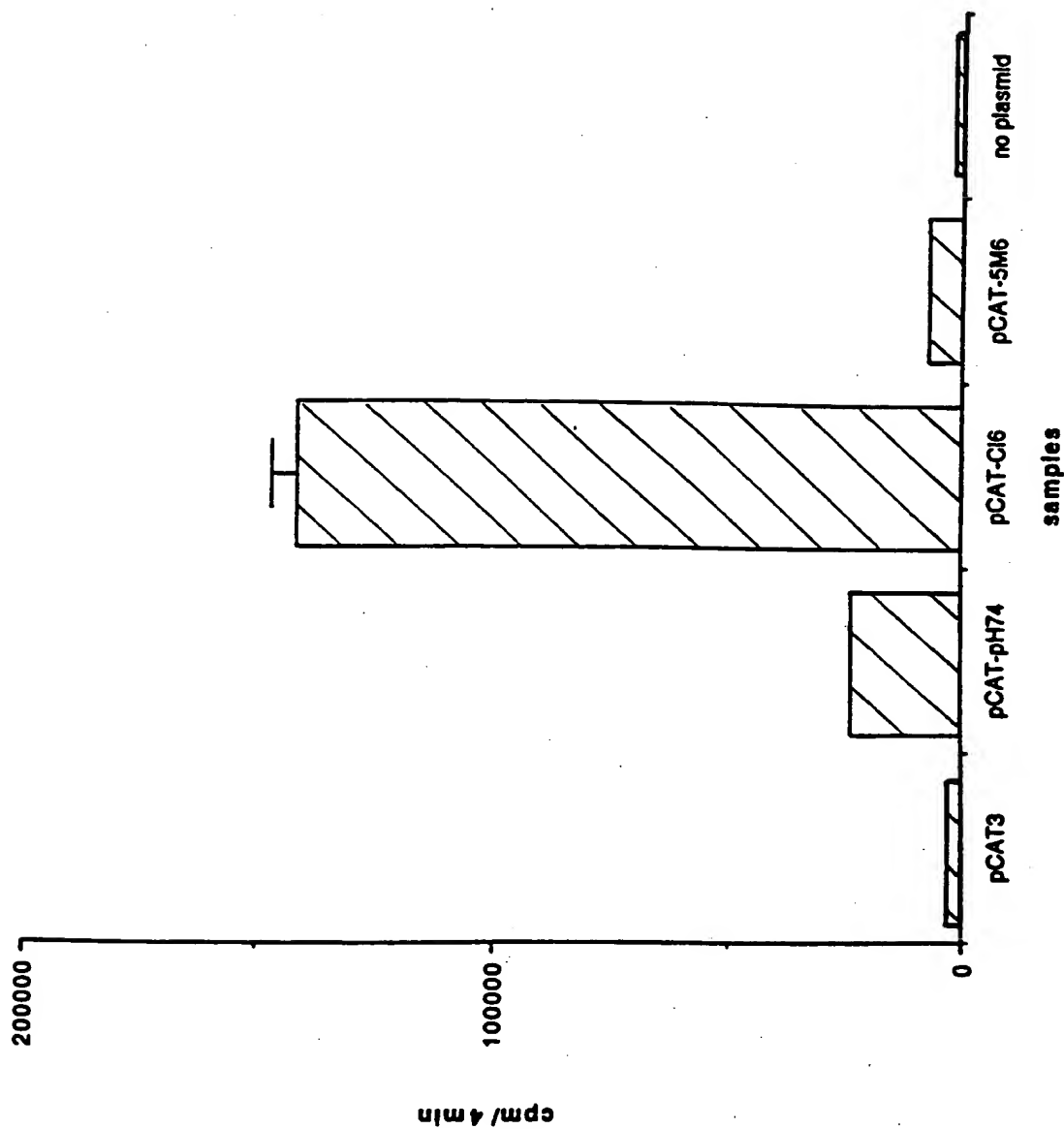


FIG12

ATGGCCCTCC CTATCATAC TTTCCTCTTT ACTGTCTCTT TACCCCTTTT GCTCTCATCT GACCCCTCTC CATCTCTCTG TACATCATCT ACTCTCCCTT
 H A L P Y H T P L P T Y L L P P P A L I A P P P C C C T T B S S P Y
 ACCAGAGATT TCTATGAGCA ACCGCGCTTC CTGGAAATAT TCATGCCCA TCTATGAGA GTTTATCTTA GGGAACTCC ACTCTCACTG CCCACACCA
 Q E P L . R T R L P G N I D A P S Y R S L S K G N B Y F T A H T H
 TATGCCCGCC AKCTCTATA ACTCTGCCAC TCTTTCATG CATGAAATA CTATATATG GACAGGAAA ATGATTAATC CTATCTTCTC TGGAGACTT
 H P R N C Y N S A T L C N H A N T H Y W T O K N I N P S C P O G L
 GAGGCCCTG TCTTTTGGAC TTACTTCAAC CATACAGTA TCTCTGATG GGTGGATT CAGCTCAGG CAGAGGAAA ACAATGAAAG GAGCAATCT
 G A T V C W T Y P T H T S N S D G G I Q G O A R E K Q V K E A I S
 CCCAACTGAC CGCGGAGCAT AGCAGCCCTA GCGCTACAA AGCACTAGT CTCTCAAAAC TACATGAAC CTTCTGTACC CATACTCGGC TGGTAGCCT
 Q L T R G H S T P S P Y K G L V L S K L H E T L R T H T R L V S L
 ATTATTAAC ACCCTCATC GGTCTCAGTA GGTCTAGCC CAAGACCTA CTATCTT N C W H C L P L H P R P Y I S I
 P N T T L T R L H E V S A Q N P T N C W H C L P L H P R P Y I S I
 CCTTTCTG AGCAATGGA CAATCTCAG ACAGAAATA ACACCACTTC CGTTTATGTA GAACTCTTG TTTCATCT GGAATGAAE CATACTCAA
 P V P E Q W N N P S T E I N Y T S V L V Q P L V S N L E I T H T S N
 ACCTCATG TGTAAATAT AGCAATAGTA TAGACAGAC CAGCTGCCA TGCATCAGGT GGTATACAC TCCACAGCA ATAGCTGCC TACCTCAGG
 L V C V K P B N T I D T T S S Q C I R W V T P P T R I V C L P B O
 AATATTTT GTCCTGCTA CCTCAGCTA TATTTCTTG AAGGCTTT CAGATCTAT GTCTCTCTC TATCTTTAG TCCCTCTAT GACCATCTAC
 I F P V C G T S A Y H C L N G S S E S H C F L S F L V P P N T I Y
 ACTGAACAG ATTATACAA TCACTGCTA CCTAGGCC ACACAAAG AGTACCAAT CTCTCTTTC TTATCAGAC AGAGTCTTA GGCAGACTAG
 T E Q D L Y N H V V P K P H N K R V P I L P P V I R A G V L G R L G
 GTATGGCAT TGCATATC ACACCTCTA CTGATCTA CTACAACTA TCTCAAGAA TAAATGTGA CATGGAACAG GTACCTACT CCTGTGTCAC
 T G I G S I T T S T Q P Y Y K L S Q E I N G D H E Q V T D S L V T
 CTTCAGAT CAATTAAT CCTTAGAC AGTAGCTTT CAATTCGA GACCTTGA CTCTTACC GCGAAAGAG GCGGAACCTG TTATTTTTA
 L O D Q L N S L A A V V L Q N R R A L D L L T A K R G G T C L P L
 GACAGAAC GCTGTATTA TGTATGCA TCCAGATTO TCACTGAA AGTAAAGA ATTGAGATC GAATACATG TAGAGGAG GAGCTCAAA
 G E R C Y Y V N Q S R I V T E K V K E I R D R I Q C R A E E L Q N
 ACAGAAAG CTGGGCTCT CTAGCCCAT GGTCTCTG GGTCTCTC TTCTTAGAC CTCTAGCAG TCTATATG TTACTCTCT TTTGACCTG
 T E R W G L L S Q W H P W V L P F L G P L A A L I L L L L P G P C
 TATTTTAA CTCTTGTTA AGTTTCTC TTCCAGATT GAGCTGTA AGCTAGAT GTCTTACAA ATGCAACCC AGATGAGTC CATGCTAAG
 I F N L L V K P V S S R I E A V K L Q H V L Q H E P Q H E S M T K
 ATCCACCTG GACCCCTGA CGGCTCTT AGCCATCT CCGATTTA TGACATGAA GGCACCTCT CCGAGGAAT CTCACTGCA CAACCTCTAC
 I H R G P L D R P A S P C S D V N D I E G T P P E E I S T A Q P L L
 TATGCCCA TTCAGGGA AGCATTA GCGCATCA GCGACCTCC CCACAGCAC TTGGTTTTC CTCTTCAGAG GCGGACTCA GAGACAGAC
 C P N S A G S S
 TAGTGGAT TGTAGGCA ACAGAAATC CTTAGCTA GCTGGAGG TGAATCAT CACCTTAA CATGGGCTT GCACTTAC TCACACCGA
 CCAATCAG AGCTACTTA ATCTAAT AGGAAAT AGAGGTAA GAATAGCA ATCATATT GCTTGAGAC ACAGCGAG GAGAAAGAT
 CGGATATA ACCAGCAT TCAAGCGGC ACGCAAT CCTTTGGT CCGCTCTT TGTATGCG CTCTGTTTC ACTTATTTC ACTTATTAA
 ACTTGCAC TGAAGAAA AAAAAAAAA
 Poly A signal
 2010

FIG13

The sequence shown in this figure is a partial sequence of the cDNA for the poly A signal. The sequence is shown in the 5' to 3' direction. The sequence is shown in the 5' to 3' direction. The sequence is shown in the 5' to 3' direction.

FIG 14

CAGCAACCCC CTCTGGTCC CTCCCATTTG TATGGGAGCT CTGTTTTCAC TCTATTTCAC TCTATTAAAT CATGCAACATG CACTTCTCTG GTCCGTGTTT 100
 TTATGGCTC AAGCTGAGCT TTGTGTGGCC ATCCACCACT GCTGTTTGCC ACCGCTGCT GACTTCCATC CCTTTGGATC CAGCAGAGT 200
 TCCGCTGTC TCCTGATCCA GCACAGGCGC CCATGTGCTC TCCCAATGG GCTAAAGCT TCGCATTTGT CCTGCACAGC TAAGTGCCTG GGTTCATCCT 300
 AATGAGCTG AACACTAGTC ACTGGGTTC ACAGGTCTCT TCCATGACCC ATGGTCTCTA ATAGAGCTAT AACACTCACT GCATGGTCCA AGATTCCATT 400
 CCTTGGATC CGTGAGACCA AGAACCCCG GTCAGAGAAC ACAGGCTTG CCACCATTTT GGAAGCAGCC CACCACCATT TTGGAAGCAG CCGGCCACTA 500
 TCTTGGAGC TCTGGGAGCA AGGACCCCG GTCACCAATTT GGTGACCCAG AAGGGACCTG AATCGCAAC CATGAAGGA TCTCCAAGC ^{gag}ATGGGAAC 600
 GTTCCCCCG AGSCAAAAT GCCCCTAGAA CGTATTTCTGG AGAATTGGGA CCATGTGAC ACTCAGAGCG TAAGAAAGAA ACCATTATA TTCTTCTGCA M G N 3
 V P P E A K M P L E R I L E N W D Q C D T Q T L R K K R F I F F C S 700
 GTACGCGCTG GCCACAATAT CCTCTTCAAG GGAGAGAAC CTGCTTCTCT GAGGAAGTA TAAATTATA CATCATCTTA CAGCTAGACC TCTTCTGTAG 37
 T A W P Q Y P L Q G R E T W L P E G S I N Y N I I L Q L D L F C R 800
 AAAGGAGGC AATGGAGTG AAGTGCCATA TGTGCAAACT TTCTTTTCAT TAAGAGACAA CTCACAATTA TGTAAAGT GTGGTTTATG CCTACAGGA 900
 K E G K W S E V P Y V Q T F F S L R D N S Q L C K K C G L C P T G 103
 AGCCCTCAGA GTCCACCTCC CTACCCGAGC GTCCCTTCCC CGACTCTTTC CTCACCTAAT AAGGACCCCT CTTTAACCCA AACGGTCCA AAGGAGTAG 1000
 S P Q S P P P Y P S V P S T P S S T N K D P P L T Q T V Q K E I D
 ACAAGGGT AAACAATGAA CCAAGAGTG CCAATATTC CCGATTATG CCCCTCAAG CAGTGAGAGG AGGAGAATTC GGCCAGCCA GAGTGCTGT 137
 K G V N N E P K S A N I P R L C P L Q A V R G G E F G P A R V P V 1100
 ACCTTTTCT CTCTCAGACT TAAAGCAAT TAAATAGAC CTAGTAAAT TCTCAGATAA CCCTGACCGC TATATTGATG TTTTACAGG GTTAGGACAA 170
 P F S L S D L K Q I K I D L G K F S D N P D G Y I D V L Q G L G Q 1200
 TCCTTTGATC TGACATGGAG AGATATATG TTACTACTAA ATCAGACACT AACCCCAAT GAGAGAAGTG CCGCTGTAAC TGACGCCGA GAGTTGGCG 203
 S F D L T W R D I M L L L N Q T L T P N E R S A A V T A A R E F G D 1300
 ATCTTTGCTA TCTCAGTCAG GCCAACAATA GGATGACAC AGAGGAAGA ACACTCCCA CAGGCCAGCA GGCAGTTCCC AGTGTAGACC CTCATTGGGA 237
 L W Y L S Q A N N R M T T E E R T T P T G Q Q A V P S V D P H W D 1400
 CACAGAATCA GAACATGGAG ATTGGTGCCA CAAACATTTG CTAACTTGG TGCTAGAAG ACTGAGGAA ACTAGAAGA AGCCTATGAA TTACTCAATG 270
 T E S E H G D W C H K H L L T C V L E G L R K T R K K P M N Y S M 1500
 ATGTCCACTA TAACACAGG AAAGGAAGAA AATCTTACTG CTTTCTGGA CAGACTAAGG GAGGCATGA GGAAGCATAC CTCCTGTGCA CCTGACTCTA 303
 M S T I T Q G K E E N L T A F L D B L R E A L R K H T S L S P D S I 1600
 TTGAAGCCA ACTAATCTTA AAGGATAGT TTATCACTCA GTCAGCTGCA GACATTAGAA AAAACTTCA AAGTCCGTC TTAGGCTCGG AACAAACTT 337
 E G Q L I L K D K F I T Q S A A D I R K K L Q K S V L G S E Q N L 1700
 AGAAACCTTA TTGAACCTGG CAACCTCGGT TTTTATAT AGAGATCAGG AGGACAGGC AGNATGGGAC AAATGGGATA AAAAAGG GGCACCGCT 370
 E T L L N L A T S V F Y N R D Q E E Q A E W D K W D K K R A T A 1800
 TTAGTCATGG CCTCAGGCA AGCGACTTT GGAGGCTCTG GAAAGGGA AAGCTGGCA AATAGGAAGC CTAATAGGCG TTGCTTCCAG TGCGGTCTAC 403
 L V M A L R Q A D F G G S G K G K S W/.A N R K P N R A C F Q C G L Q 1900
 AAGGACACTT TAAAAAGAT TGTCGAATA GAATAAGCC GCCCCTTGT CCATGCCCT TACGTCAAG GAATCACTGG AAGGCCACT GCCCAGGG 437
 G H F K K D C P N R N K P P C/R P C P L R Q G N/.H W K A H C P R G 2000
 ATCAAGATAC TCTGAGTCAG AAGCCATTA CCAGTATATC CAGCAGCAGG ACTGA S R Y S E S E A I N Q M I Q Q Q D 470
 S R Y S E S E A I N Q M I Q Q Q D 487

FIG 15

100 GGACCGTAG TATGGGTAA TCCCTCCGG GAACCAAGC CCAGTACTC AGAAGAAGAA ATAGAATGGG GAACCTCAGC AGGACATGGT TTCTCCCT
 34 GIPVVWGNPLRETKPEYSEEEIEWGTSRGHGFLPS
 200 CAGGATGGCT AGGCACTGAA GAAGGAAGAA TACTTTTGT GGCAGCTAC CAATGGAAT TACTTAAAC CTTTCAGCAA ACCTTCACT TAGGCATTGA
 67 G W L A T E E G K I L L L A A N Q W K L L K T L Q Q T P E L G I D
 300 TAGCACCAT CAGATAGCCA AATCATATT TACTGACCA GGCCTTTCA AACTATCA GCGATAGTC AGGCTCTG AGTGTGCCA AGAATATAT
 100 S T E Q I A K S L F T G P G L P K T I K Q I V R A E E V E Q R N M
 400 CCGCTGCTT ATCGCAAGC TCCTTCAGGA GAACAAAGAA CAGGCAATTA CCCAAGAGAA GACTGGCAAC TAGATTTTAT CCACATGCCA AATCAGG
 134 P L P Y R Q A P S G E Q R T G N Y P R E D W Q L D F I H M P K S Q G
 500 GATTCAGTG TCTACTAATC TGGTAGATA CTTTCACTGG TTGGCAGAG GCTTCCCT GTAGGACAGA AAGTTCCTA GAGGTATTA AGGCACTAGT
 167 F Q C L L V W V D T F T G W A E A P P C R T E K P Q E V I K A L V
 600 TCATCAAGTA ATCCCAAT TCAGACTTC CTGAGGCTTA CAGAGTGACA ATGCTCTGC TTCAAGGCC ACAGTAACC AGGAGTATC CCAGGCTTA
 200 H E V I P R P G L P G L Q S D N G P A F K A T V T Q G V S Q A L
 700 GGTATAGAT ATCACTTACA CTGCACCTAG AGGCACAAAT CCTCAGGAA GGTGAGAA ATGATACAC TCAAGGACA TCTAAGCAG CTACCCAGG
 234 G I E Y H L H C T . R P Q S S G K V E K M K T L K R H L N K L T Q E
 800 AAACCACT GGCATGGT GCTCTGTGT CTATAGCTT ACTAAGATC CAATCTCT CCAAGGCC AGGACTAGC GCATACAGAA TGCTGTATGG
 267 T H L A W S A L L S I A L L R I Q N S P Q K A G L S P Y R M L Y G
 900 ACGTCTTC CTACCAATG ACCTCTGCT TGACCAAGAG ATGGCCAACT TAGTGCAGA CATCACTCC TTAGCCAAAT ATCAACAAT TCTTAAACA
 300 R S F L T N D L L L D Q E M A N L V A D I T S L A E Y Q Q V L K T
 1000 TTACAGGAG CTTGTCCCG AGAGGAGGA AAGAAATAT TCCACCTGG TGTATGTA TTAGTCAAGT CCGTCCCTC TAATCCCA TCCCTAGACA
 334 L Q G A C P R E E G K E I F H P G V M V L V K S L P S M S P S L D T
 1100 CATCTGGG AGGACCTAC CCAATATT TATCTATCC AACTCGGT AAGTGGCTG GAGTGGATC TTGATACAT CACTCTGAA TCAACCTG
 367 S W G G P Y P V I L S I P T A V K V A G V E S W I H H T R I K P W
 1197 GATACTGGG AAGAACCCG AAATCCAGG GGACACGCT AGCTATTCT TTGAACCTT AGAGATCTG TCCTGCTCT TCAGCAACA ACCGTGA
 398 I L P K E P E N P G D N A S Y P F E P L E D L C L L P K Q Q P

K Y T N Q H

1719